Supplementary Data

PaintOmics 3: a web resource for the pathway analysis and visualization of multi-omics data

Rafael Hernández-de-Diego, Sonia Tarazona, Carlos Martínez-Mira, Leandro Balzano-Nogueira, Pedro Furió-Tarí, Georgios J Pappas Jr, and Ana Conesa

MATERIALS AND METHODS

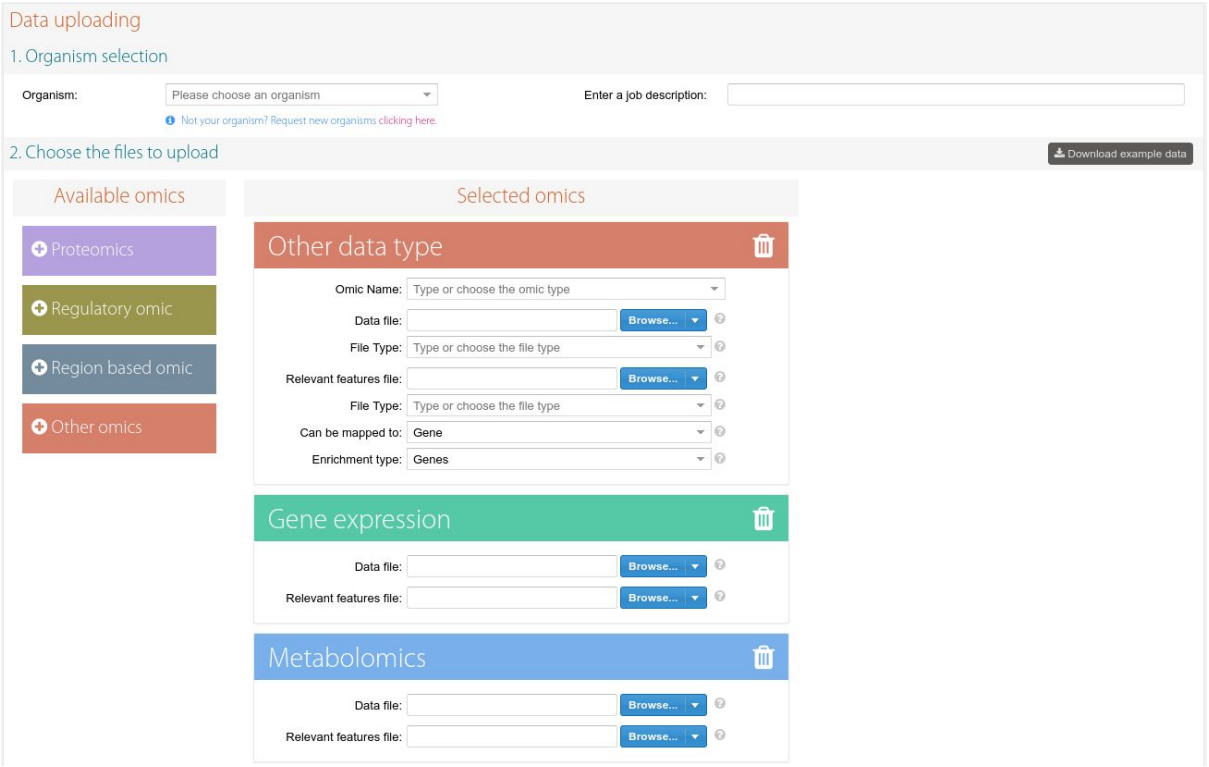
The PaintOmics 3 architecture

Supplementary Table 1. List of open-source resources used in PaintOmics 3.

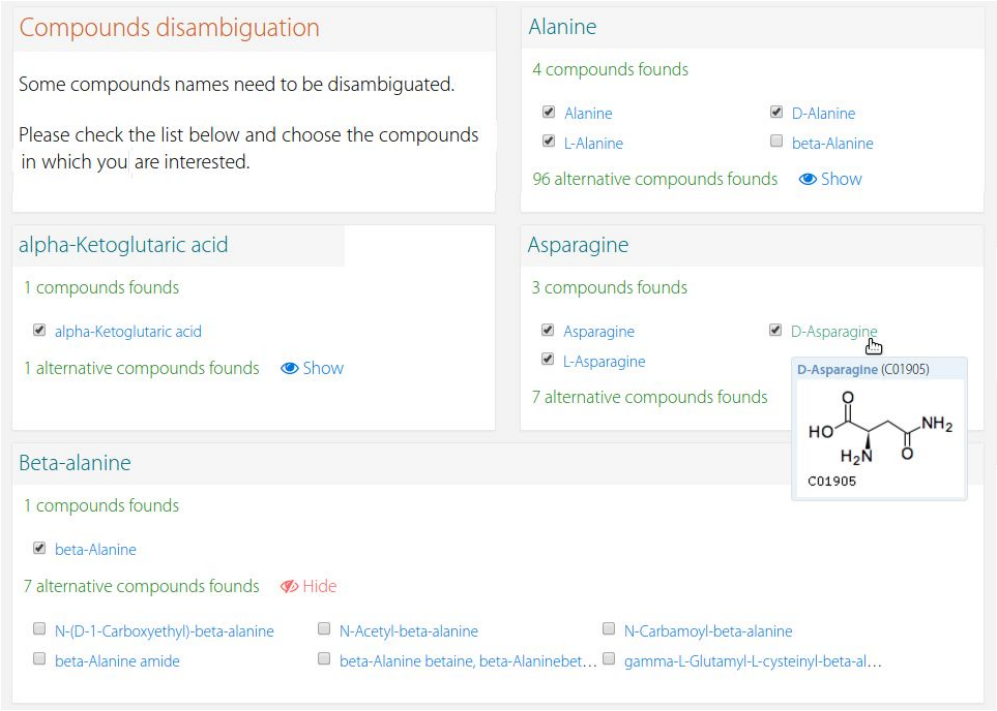
Resource name	Reference
Python 2.7	https://www.python.org/
R	https://www.r-project.org/
MongoDB	https://www.mongodb.com/
ExtJS 4.2.1	https://www.sencha.com/products/extjs/
Font awesome	https://fontawesome.com/
jQuery 3.1.0	https://jquery.com/
jQuery UI	https://jqueryui.com/
clustefck	https://harthur.github.io/clusterfck/
Dragula	https://bevacqua.github.io/dragula/
Highcharts	https://www.highcharts.com/
Linkurious	https://linkurio.us/
Odometer	http://github.hubspot.com/odometer/

svgjs	http://svgjs.com/
Tooltipster	http://iamceege.github.io/tooltipster/
randomColor	https://github.com/davidmerfield/randomColor/

Input data

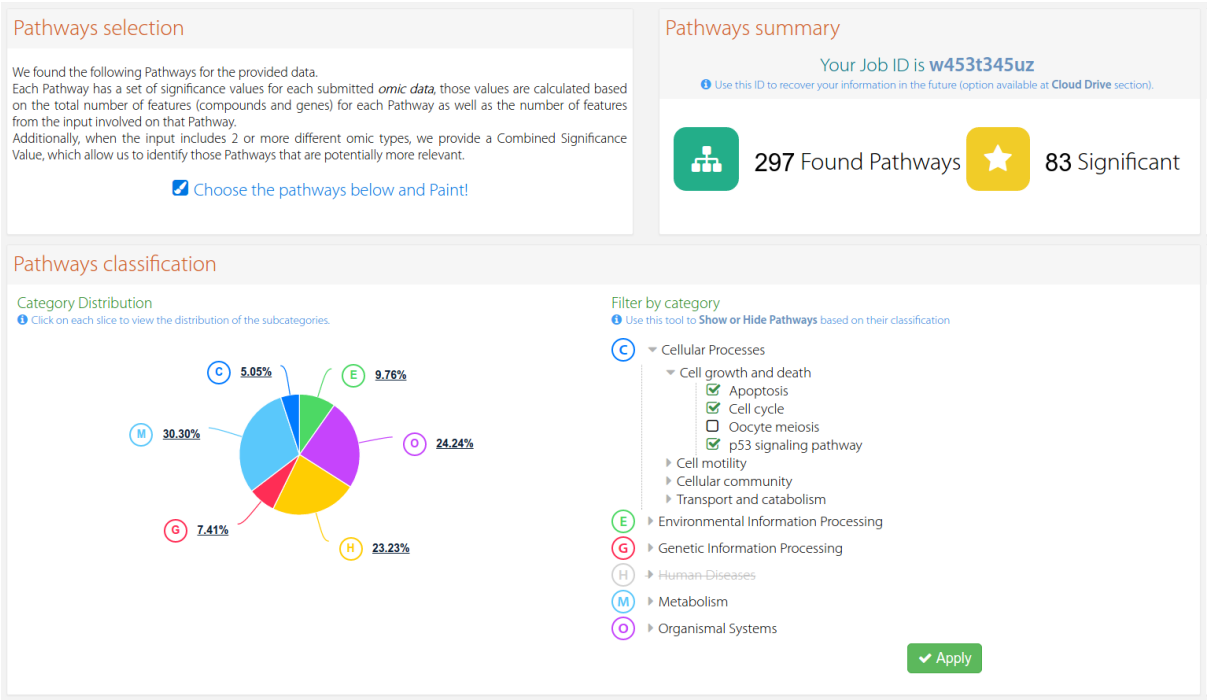


Supplementary Figure 1. PaintOmics 3 input data form.



Supplementary Figure 2. Matching metabolite names provided by the user to metabolite names in KEGG pathways in PaintOmics 3.

Hierarchical classification for KEGG pathways

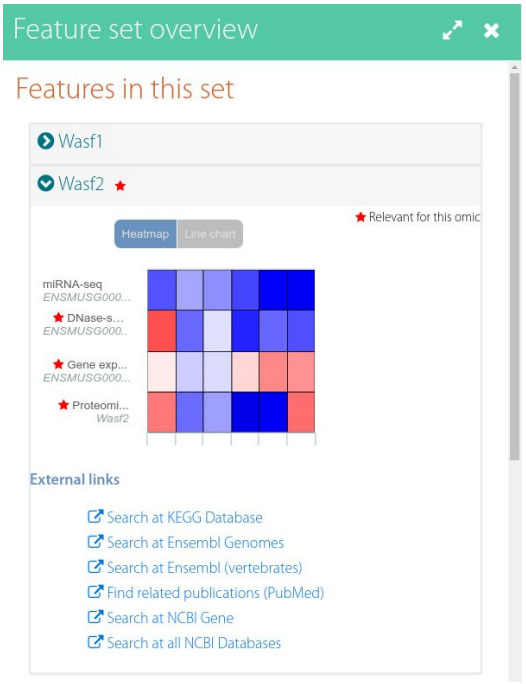


Supplementary Figure 3. Pathway classification in PaintOmics 3. Using the filtering tools, we can exclude from the results pathways that are not interesting for our specific study.

Multi-omic pathway-based visualization

The pathway visualization main panel is interactive. By clicking on any of the pathway features a floating window appears with complete multi-omics data in the form of a heatmap and offering cross-references to public databases. By clicking on *Show details*, the *Feature Set Overview* panel (Supplementary Figure 4) also shows this detailed information. This allows, for example, to readily access additional regulatory data such as the sets and values of microRNAs or transcription factors associated to the gene. Moreover, KEGG frequently includes more than one gene in the same feature box, that might represent paralogues or protein complexes. The feature floating windows indicates this multiplicity and allows navigating through the different elements of the feature box.

Finally, *Settings* button allows the user to control the appearance of the painted diagram, adjusting colors, scales and selecting the omic type to be displayed.



Supplementary Figure 4. Additional functionalities of interactive pathway visualization in PaintOmics 3.

USE CASE

Cacchiarelli's data

The Cacchiarelli' study (1) includes transcriptomics (RNA-seq and small RNA-seq), methylation (RRBS-seq) and region-based histone modification (H3K4me3 ChIP-seq) data taken at different time points after reprogramming. To build our example we selected time

points with data for all omic types: hiF-T (human inducible fibroblasts-like cells), day 5, day 10, day 24, day 24* (cells reprogrammed for 20 days in doxycycline followed by 4 days without doxycycline), and hIPSC-T (human Induced Pluripotent Stem Cells).

Data pre-processing

RNA-seq and miRNA-seq. Non-expressed genes in all conditions were removed and TMM (Trimmed Mean of M-values) normalization (2) was applied.

Methyl-seq. CpG sites with missing values were removed. Remaining CpG sites were associated with RGmatch tool (3) to the closest genes. For each gene, we selected CpG sites in its promoter region (1 kb upstream the transcription start site) and averaged their methylation values.

ChIP-seq. We obtained consensus peaks across all six conditions included in the dataset by merging overlapping regions with at least one shared nucleotide and computing the widest intersecting range. For each time point, the presence/absence of the consensus region was coded by 1 or 0 respectively.

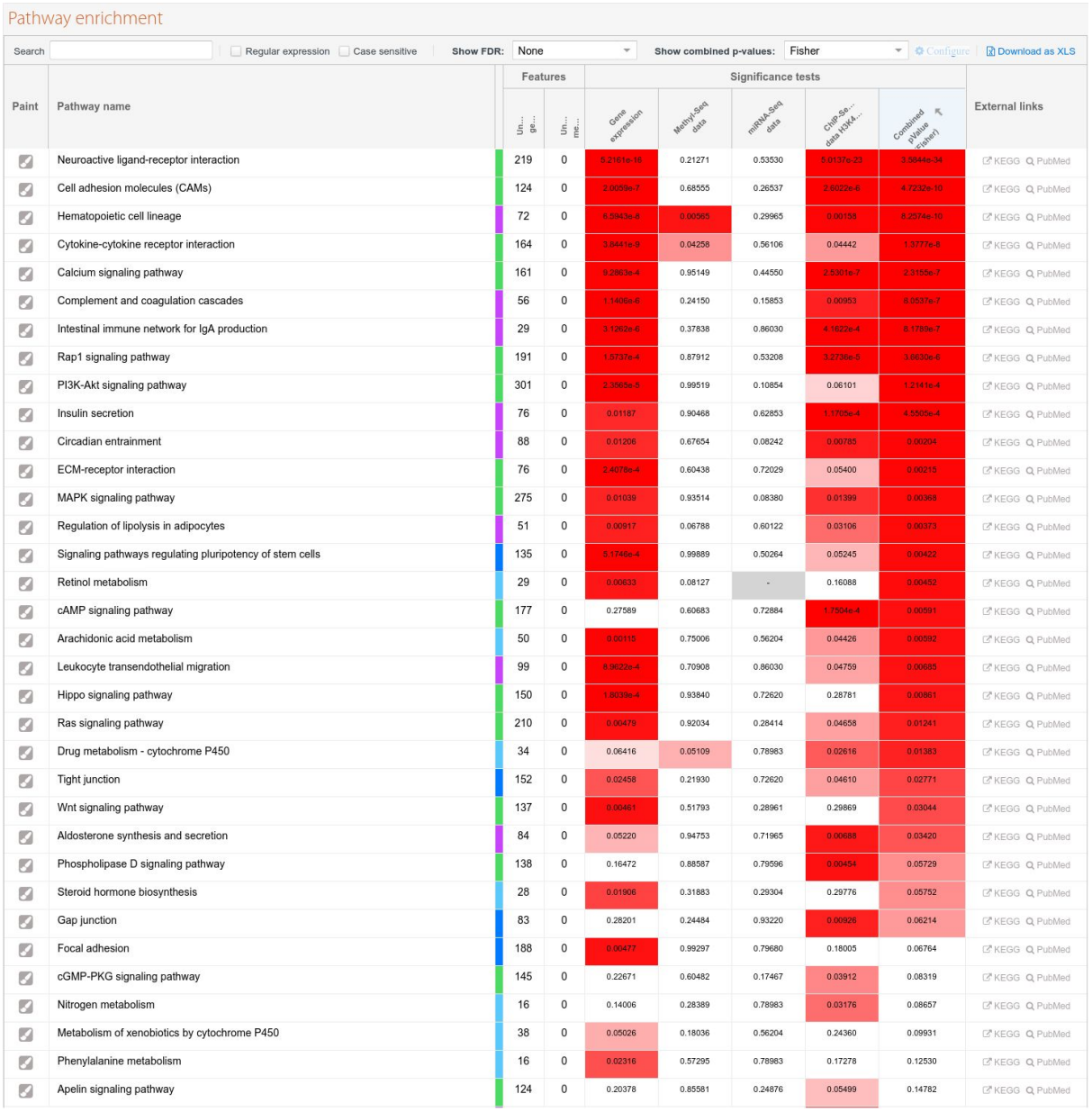
Paintomics input data

The hiF-T sample was taken as the initial time point to calculate a log2-fold change for the rest of time points, replacing zeros with a convenient value to avoid indeterminations.

Exceptionally for ChIP-seq data, "fold-change" value was defined for each time point as 1 if the peak was found at that time point but not in the reference, 0.5 if found in both, 0 if not found in any of them, and -1 if only found at the initial time point.

Relevant features were considered those with an absolute log2-fold change higher than 5 at any of the time points for RNA-seq and ChIP-seq, and higher than 10 for methylation. We obtained 15% relevant genes for RNA-seq, 58% relevant miRNAs and 12.4% relevant genes for methylation. For ChIP-seq, we selected as relevant peaks those including a "fold-change" value of 1 or -1 in at least 3 time points, which resulted in 24% of relevant peaks. The miRNA-gene association file for human was obtained from miRWalk2.0 (4) and PaintOmics 3 miRNA matching tool was set to only keep pairs with correlation lower than -0.3. PaintOmics 3 processing of H3K4me3 at the Region based omic tool was set to retain peaks at promoter regions defined as 1.5 kb upstream transcription start sites.

Pathway enrichment analysis



Paint	Pathway name	Un... ge...	Un... me...	Gene expression	Methyl-Seq data	miRNA-Seq data	ChIP-Se... data H3K4...	Combined p-Value (Fisher)	External links
	Neuroactive ligand-receptor interaction	219	0	5.2161e-18	0.21271	0.53530	5.0137e-23	3.5844e-34	KEGG PubMed
	Cell adhesion molecules (CAMs)	124	0	2.0059e-7	0.68555	0.26537	2.6022e-6	4.7232e-10	KEGG PubMed
	Hematopoietic cell lineage	72	0	6.5943e-8	0.00565	0.29965	0.00158	8.2574e-10	KEGG PubMed
	Cytokine-cytokine receptor interaction	164	0	3.8441e-9	0.04258	0.56106	0.04442	1.3777e-8	KEGG PubMed
	Calcium signaling pathway	161	0	9.2863e-4	0.95149	0.44550	2.5301e-7	2.3155e-7	KEGG PubMed
	Complement and coagulation cascades	56	0	1.1406e-6	0.24150	0.15853	0.00953	8.0537e-7	KEGG PubMed
	Intestinal immune network for IgA production	29	0	3.1262e-6	0.37838	0.86030	4.1622e-4	8.1789e-7	KEGG PubMed
	Rap1 signaling pathway	191	0	1.5737e-4	0.87912	0.53208	3.2736e-5	3.6630e-6	KEGG PubMed
	PI3K-Akt signaling pathway	301	0	2.3665e-5	0.99519	0.10854	0.06101	1.2141e-4	KEGG PubMed
	Insulin secretion	76	0	0.01187	0.90468	0.62853	1.1705e-4	4.5505e-4	KEGG PubMed
	Circadian entrainment	88	0	0.01206	0.67654	0.08242	0.00785	0.00204	KEGG PubMed
	ECM-receptor interaction	76	0	2.4078e-4	0.60438	0.72029	0.05400	0.00215	KEGG PubMed
	MAPK signaling pathway	275	0	0.01039	0.93514	0.08380	0.01399	0.00368	KEGG PubMed
	Regulation of lipolysis in adipocytes	51	0	0.00917	0.06788	0.60122	0.03106	0.00373	KEGG PubMed
	Signaling pathways regulating pluripotency of stem cells	135	0	5.1746e-4	0.99889	0.50264	0.05245	0.00422	KEGG PubMed
	Retinol metabolism	29	0	0.00633	0.08127	-	0.16088	0.00452	KEGG PubMed
	cAMP signaling pathway	177	0	0.27589	0.60683	0.72884	1.7504e-4	0.00591	KEGG PubMed
	Arachidonic acid metabolism	50	0	0.00115	0.75006	0.56204	0.04426	0.00592	KEGG PubMed
	Leukocyte transendothelial migration	99	0	8.9622e-4	0.70908	0.86030	0.04759	0.00685	KEGG PubMed
	Hippo signaling pathway	150	0	1.8039e-4	0.93840	0.72620	0.28781	0.00861	KEGG PubMed
	Ras signaling pathway	210	0	0.00479	0.92034	0.28414	0.04658	0.01241	KEGG PubMed
	Drug metabolism - cytochrome P450	34	0	0.06416	0.05109	0.78983	0.02616	0.01383	KEGG PubMed
	Tight junction	152	0	0.02458	0.21930	0.72620	0.04610	0.02771	KEGG PubMed
	Wnt signaling pathway	137	0	0.00461	0.51793	0.28961	0.29869	0.03044	KEGG PubMed
	Aldosterone synthesis and secretion	84	0	0.05220	0.94753	0.71965	0.00688	0.03420	KEGG PubMed
	Phospholipase D signaling pathway	138	0	0.16472	0.88587	0.79596	0.00454	0.05729	KEGG PubMed
	Steroid hormone biosynthesis	28	0	0.01906	0.31883	0.29304	0.29776	0.05752	KEGG PubMed
	Gap junction	83	0	0.28201	0.24484	0.93220	0.00926	0.06214	KEGG PubMed
	Focal adhesion	188	0	0.00477	0.99297	0.79680	0.18005	0.06764	KEGG PubMed
	cGMP-PKG signaling pathway	145	0	0.22671	0.60482	0.17467	0.03912	0.08319	KEGG PubMed
	Nitrogen metabolism	16	0	0.14006	0.28389	0.78983	0.03176	0.08657	KEGG PubMed
	Metabolism of xenobiotics by cytochrome P450	38	0	0.05026	0.18036	0.56204	0.24360	0.09931	KEGG PubMed
	Phenylalanine metabolism	16	0	0.02316	0.57295	0.78983	0.17278	0.12530	KEGG PubMed
	Apelin signaling pathway	124	0	0.20378	0.85581	0.24876	0.05499	0.14782	KEGG PubMed

Figure 5. Results for the PaintOmics pathway enrichment analysis for Cacchiarelli's data. A total of 323 pathways were initially reported, but some pathway categories were excluded as they were irrelevant for the present study. Thus, 199 pathways were tested for enrichment and 25 of them were considered to be significantly enriched by the combination of all the omics (p-value < 0.05). The enriched pathways are ordered by combined p-value. Upper positions correspond to the most significant pathways. A color scale is used to highlight the level of enrichment for each pathway where the higher the intensity of red, the higher the significance. Gray cells indicate that the corresponding omic data type is not present in the pathway.

REFERENCES

1. Cacchiarelli, D., Trapnell, C., Ziller, M. J., Soumillon, M., Cesana, M., Karnik, R., Donaghey, J., Smith, Z. D., Ratanasirintrawoot, S., Zhang, X., et al. (2015) Integrative analyses of human reprogramming reveal dynamic nature of induced pluripotency. Cell, 162(2), 412–424.
2. Robinson, M. D. and Oshlack, A. (2010) A scaling normalization method for differential expression analysis of RNA-seq data. Genome biology, 11(3), R25.
3. Furio-Tari, P., Conesa, A., and Tarazona, S. (2016) RGmatch: matching genomic regions to proximal genes in omics data integration. BMC Bioinformatics, 17(Suppl 15), 427.
4. Dweep, H. and Gretz, N. (2015) miRWalk2.0: a comprehensive atlas of microRNA-target interactions. Nat. Methods, 12(8), 697.